ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut

CENTER DISCUSSION PAPER NO. 334

THE GROWTH AND DECLINE OF CHINESE FAMILY CLAN

John C. H. Fei and Ts'ui-jung Liu*

January 1980

The Growth and Decline of Chinese FAmily Clan

Introduction

One well recognized cultural trait of traditional China was
the extented family system, i.e., the family clan (chia-tzu 家族). A
family clan may contain, at any moment of time, well over thousands of
members performing many educational, ceremonial, economic, and social
security oriented functions that can be executed efficiently only when the
size of the clan is large enough.[1] While in principle, a particular clan
begins with a single male ancestor, the starter, a central dynamic phenomenon
is that the membership will undergo an expansion process that may take
place over several centuries. A basic instrument that a clan uses
to establish the clan identity is the genealogy that provides a record
of this dynamic process. While thousands of volumes of Chinese genealogies
constitute a primary data source for scholarly research in many areas,[2] we
shall in this paper make use of the genealogical data for one particular
purpose, namely, to trace the rules of growth and decline of a family clan.

The growth of the membership of a family clan is first of all a
demographic phenomenon following the rules (such as the age-specific birth
and death rates) established in modern demography. However, in applying
these rules to the study of Chinese genealogies, two additional sources of
complexity (or dimensions) arise. The first dimension is the necessity for
an explicit identification of the generation. The growth of a Chinese
family clan is a male oriented regeneration process in which great

significance is attached to the hierarchy of status based on the generational identity (i.e., a grand-uncle has a higher status than an uncle) as well as the seniority based on age. While the age distribution is the focal point of analysis in modern demography, the generation dimension that is usually neglected must be introduced explicitly.

The second dimension is that the growth of the family clan has a life cycle characterized by the existence of a growing phase followed by a declining phase. Under this situation the rules of growth, especially in the declining phase, are only partially demographic in nature.[3] Our purpose is to identify the various phases of the life cycle and to investigate the rules of growth in the different phases.

We shall begin with the observed male birth schedules to construct what will be called the hierarchy matrix that describes the clan membership at any moment of time stratified by ages and generations (Section 1). By investigating the time path of the clan population, we shall account for the exogenous influences and identify three phases of the life cycle of the clan to be called the generative phase, the transition phase, and the declining phase, both empirically and heuristically (Section 2). We shall then concentrate on the shape (i.e., the pattern) of the generation birth schedules to establish certain "stylized facts" empirically (Section 3). A theoretical model (i.e., a "Single Ancestor Model") will be presented to explain these stylized facts (Section 4). This model will be revised (i.e., replaced by a "Multiple Ancestor Model") in order to accommodate the incompleteness of genealogical data for the earlier generations (Section 5). In the context of this model, the theory on the size of the family clan will be tested with our data to mark-off the generative phase from the transition phase (Section 6). The disintegration of the family clan

during the transition phase will be examined from the nature of the birth range and the structure of the hierarchy matrix theoretically and empirically (Section 7). Finally, we shall speculate on the cause of the disintegration as traced to population pressure on land (Section 8).

In the text, we shall present the theory in non-technical terms. The technical and formal derivation of the dynamic relations will be presented in the appendices. For empirical study, we shall make use of the genealogy of one particular Hsü clan in Hsiao-shan, Chekiang. The conclusion of our paper can be substantiated or refuted by future studies following the methodology suggested in this paper.

Section 1: The Generation Birth Schedule and Hierarchy Matrix

The generation birth schedules for the Hsü clan males are listed in Table 1. The numbers in each row form a generation birth schedule showing the number of male births in each five-year interval with the mid-year indicated on the top of the table. While the total number of male births in each generation is indicated at the right hand side margin, the total number of male births in each five-year interval is indicated in row (10). This row is shown by the curve labeled "Observed total births" in Diagram 1. If the generation birth schedules are denoted by b^0, b^1, b^2,...the observed total birth can be denoted by $b^0 + b^1 + b^2 + \ldots$. The birth schedules are the basic information which can be abstracted from the genealogies that recorded the birth dates of the members stratified by generations. The data of birth schedules of the Hsü clan listed in Table 1 covered nine generations (i.e., the eleventh through the nineteenth) ranging over nearly 250 years (1665-1910).[4]

From the birth schedules of Table 1, we can derive a hierarchy matrix of the clan for any particular year. As an illustration, the hierarchy matrix for the year 1790 is shown in Table 2. To derive this matrix, assuming the male longevity is 80 years, the segment of the birth schedules from 1710 to 1790 with a age range of 80 years (as blocked out in Table 1) is reproduced in Table 2. Notice that a total of seven consecutive generations (i.e., the 10th through the 16th) are listed in Table 2. This implies that seven generations coexisted in 1790. We shall investigate this issue in section 6 from a theoretical standpoint.[5]

The entries in Table 2 are multiplied by the survival rates[6] given in the row next to the bottom line to show the number of males in each generation and age group. Those numbers are contained in the parentheses of each cell. The total number of males surviving in 1790 stratified by age and generation are also listed, giving a total male population of 702. Notice that the total number of survivors declines monotonically with age, conforming to a familiar demographic pattern. The total number of male survivors for the various generations takes on a u-shaped pattern, indicating that there are relatively few descendants for the older and younger generations during any moment of time.

The hierarchy matrix portrays a two-way ordering of member status under the family clan system. A person generally has higher clan status if he belongs to an earlier generation and/or is older. For example, individuals in cell A have a higher status than those in the block ABCD (i.e., that lies in the southeast) and have a lower status than those in AFGH. For certain ceremonial purposes (e.g., the listing of names in a funeral

announcement or in ancestor worshiping), the generation ordering takes
precedence regardless of the age, while the opposite is true in other clan-
related functions (e.g., education and assignment of duty and authority).
Thus the hierarchy matrix for any year indicates not only the size of the
clan but also its status structure. It is clear that whatever functional
significance (e.g., education, ceremony, and relief, etc.) that the clan
system might have is definable mainly by the entities contained in the
hierarchy matrix.

Section 2: Clan Population Size and Exogenous Influences

The "generation conscious" genealogies usually do not provide a
record of the total population of the clan. However, we can estimate
the total male population for every year by the method just described,
as summarized in row (11) of Table 1. This now is shown by the curve
in Diagram 2 labeled "Total male population",[7] which demonstrates the growth
of the clan size (taking the size of male population as a proxy) over
time. The total male births of Diagram 1 and the total male population of
Diagram 2 together provide much of the life cycle information of the
clan.[8]

The growth of Hsu clan through two centuries (1700-1900) was
interrupted by certain major exogenous events. From the observed
total birth schedule, two major interruptions occurred between 1745 and
1765 and between 1820 and 1845. For the first interruption, the gazetteer
of Hsiao-shan recorded a major famine in 1748 such that, "Even the grass
roots and barks were exhausted as the source of food supply and people died
of eating the kuan-yin 觀音 earth which they dug up from the ground." For

the second major interruption, the gazetteer stated that in 1820, "A major drought occurred between May and July and the river was dried to the bottom to be followed by a flood brought about by typhoon, so that nearly eighty percent of the county was inundated with no hope of an autumn harvest." In fact, even minor natural calamities were faithfully reflected by minor dips in the birth schedule as indicated in Diagram 1.[9] Thus we may conclude that the genealogical data are indeed very reliable as they are sensitive to the major or even minor exogenous disturbances.

When we study the growth of the family clan from a long-run perspective, we must neglect the exogenous interruptions in order to isolate the demographic and institutional forces which delimit the phases of growth. From both the population curve and the total birth schedule, we can identify three growth phases: the generative phase (1665-1745), the transition phase (1745-1845), and the declining phase (after 1845). The declining phase clearly shows the disintegration of the family clan as both the male population and total male births decline absolutely. During the generative phase, the family clan grows with vitality as both curves show geometric rule of progression (i.e., grows with constant growth rate). During the transition phase, the growth rate decelerates as both curves show algebraic rule of growth by a long-run linear trend. In Diagram 2, the annual growth rate of the male population is shown by a dotted curve, labeled "Actual growth rate", which demonstrates a clearly declining trend during the transition phase. The demarcation of the three phases will be treated analytically in later sections.

Section 3: Inductive Evidences for Relative Frequencies

The observed generation birth schedules b^0, b^1, b^2,...of Table 1 are represented by the bar diagrams in Diagram 3. Suppose $x=(x_1,x_2,x_3,\ldots x_n)$ is any row of numbers (i.e., a row vector), we shall use $x^*=(x_1^*,x_2^*,\ldots x_n^*)$ $=(x_1/s, x_2/s,\ldots x_n/s)$ where $s=x_1+x_2+\ldots+x_n$ to denote a "normalized pattern of x". Since $x_1^*+x_2^*+\ldots+x_n^*=1$, x^* represents a "system of percentage". The normalization of b^0, b^1, b^2,... leads to b^{*0}, b^{*1}, b^{*2},... which are represented by the solid curves in Diagram 3. A b^{*t} represents the percentage distribution of male birth through time of the t-th generation. Obviously the diagram (b^t) and the solid curve (b^{*t}) have the same shape or "pattern" which is now the focus of our attention. From the observed b^t or b^{*t} (i.e., from the inductive evidences) we shall first establish certain "stylized facts" in this section.

First of all, the generation birth schedules are bell-shaped. For each b^{*t}, we can compute a mean value, e.g., $Y_0=1721.1$, $Y_1=1752.7$, and $Y_2=1784.3$ for the 11th, 12th and 13th generations as marked on the horizontal axis. The number Y_i is the mean birth year of the i-th generation. A generation birth schedule takes on a maximum value near the mean birth year and declines with larger deviations from the mean year in both directions. Thus there are relatively few very old or very young male decendants for any generation. On a closer examination we find that the shape of the bell becomes more regular (smoother and more symmetrical) for later generations. Thus the first stylized fact is that the sequence b^{*0}, b^{*1}, b^{*2},... is asymptotically bell shaped.

Secondly, it is obvious that the mean birth year gets larger for later generations (i.e., $Y_0 < Y_1 < Y_2 \ldots$). However, on a closer examination, we find the mean gap (i.e., the difference between two consecutive means, $Y_{i+1} - Y_i$) such as $Y_1 - Y_0 = 1752.7 - 1721.1 = 31.6$ and $Y_3 - Y_2 = 1784.3 - 1752.7 = 31.4$ appears to take on a constant value of slightly more than 31 years. Thus the second stylized fact is that the sequence b^{*0}, b^{*1}, b^{*2},... exhibits a constant mean gap of 31 plus years.

Finally, we find that the birth range (i.e., the age difference between the oldest and youngest male) of a generation increases for later generations. Thus from Table 1, the birth ranges are 110 (=1775-1665), 125 (=1810-1685), 135 (1855-1720) years for the 11th, 12th, and 13th generations. This diminishing concentration tendency can also be seen from the fact that the b^{*t} "bells" are getting less tall and wider for later generations. Thus the third stylized fact is that the sequence b^{*0}, b^{*1}, b^{*2},... exhibits a diminishing concentration tendency, a fact that can be verified by computing the standard deviations of b^{*t}.

The three stylized facts (asymptotically bell shaped, constant mean gap and diminishing concentration), identified purely from inductive evidences, provide guidance in our search for a theory that explains family clan growth. Any theory that is adequate must explain these stylized facts. The "shape" of the generation birth schedules is a crucial matter. For example, the "structure" of the hierarchy matrix (Table 2) clearly depends upon the shape of b^{*t} (e.g., the extent to which the generation birth schedules overlap). We shall sketch the outline of such a theory in non-technical terms in the following sections -- leaving all rigorous proofs of theorems in Appendix I.

Section 4: The Single Ancestor Model

To begin with, it is natural to postulate the existence of single male ancestor, denoted by E^0, as the starter of the family clan. E^0 will generate a "wave" of sons over time according to the male fertility schedule $U^1 = (u_0^1, u_1^1, u_2^1, \ldots u_8^1) = (.035, .282, .377, .436, .396, .253, .143, .072, .027)$. Thus u_i^1 (i=0,1,2,...8) is the probability that a typical father will generate (i.e., his spouse will give birth to) a son (i.e., a male descendant) by the time his age reaches the mid-point of the i-th age group in (15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59).[10] Thus the male ancestor E^0 generates a theoretically expected birth schedule of the next (i.e., the first) generation E^1 which is precisely U^1 (i.e., $E^1 = U^1$).

Now if every male in E^1 generates a "wave" of sons according to U^1, the sum of all "waves" becomes $E^2 = (u_0^2, u_1^2, \ldots)$, the expected birth schedule of the second generation. Thus u_i^2 (i=0,1,2,...) is the probability that a grandson is expected to be born to the ancestor E^0 in the i-th year. Recursively then the male ancestor will generate a sequence of theoretically <u>expected birth schedules</u> E^0, E^1, E^2,... for all future generations in the single ancestor model.

Since the male fertility schedule (U^1) is the only information needed to generate the entire expected sequence, we shall now define five "indices" (to be denoted by s_1, τ, n, μ_1, and σ_1^2) which describe its demographic characteristics. First, the sum of all elements of U^1 is $s_1=2.02$ which is the total number of sons that are expected to be born to a typical father during his lifetime. The magnitude of s_1 obviously governs the rapidity of growth of the clan in the long run.

Next, U^1 specifies that a male before the age of 15 cannot generate a son. Thus $\tau=3$ (or $3 \times 5 = 15$ years) is the length of the non-productive period. Similarly, $n=9$ (or $5 \times 9 = 45$ years) is the productive period, or the maximum age difference between sons of the same father. While τ governs the mean gap between E^i and E^{i+1}, the birth range of E^i is determined by the length of the productive period. In fact, the birth range of E^t is $t(n-1)+1$, so that, by the time of $t=72$, the maximum age difference between the oldest and the youngest male descendants of "Confucious" in the 72-th generation could be as high as $72(45-1)+1 = 3169$ years!

From the normalized U^1, to be denoted by $U^{*1} = (u_0^{*1}, u_1^{*1}, \ldots u_{n-1}^{*1})$ $= (u_0^1/s_1, u_1^1/s_1, \ldots u_{n-1}^1/s_1)$, we can define

i) $\underline{\text{a mean of } U^1}$: $\mu_1 = 0u_0^{*1} + 1u_1^{*1} + 2u_2^{*1} + \ldots + (n-1)u_{n-1}^{*1}$ ($\mu_1 = 3.32$)

ii) $\underline{\text{a variance of } U^1}$: $\sigma_1^2 = \sum_{i=0}^{n-1} u_i^{*1}(i - \mu_1)^2$ ($\sigma_1^2 = 2.898$)

Thus $\tau + \mu_1 = 3.32 + 3 = 6.32$ (or $5 \times 6.32 = 31.6$ years) is the average fatherhood age of a typical father. Imagine that the "age of father" is recorded on the birth certificate of every male child, then $\tau + \mu_1$ is the average "age of father" computed from a large number of such certificates. The square root of σ_1^2 is the standard deviation of U^{*1}, an index of dispersion of the male fertility schedule. Thus when the male fertility schedule is given, five indices ($s_1=2.02$, $\tau=3$, $n=9$, $\mu_1=3.32$, $\sigma_1^2=2.898$) are conceptually defined.

Let E^{*i} ($i=1,2,3,\ldots$) be the (normalized) relative frequency of the expected birth schedules E^i ($i=1,2,3,\ldots$). For convenience, let us think

in terms of an <u>ancestor calendar</u> (instead of the Christian calendar) by taking the birth year of E^0 (the ancestor) as the 0-th year. Then we have the following theorem (proved as corollary 1 in Appendix I):

<u>Basic Theorem</u>:   <u>The sequence of relative frequencies of the expected generation</u>

<u>birth schedules is asymptotically normally distributed</u>, i.e.,

<u>for large t</u>, F^{*t} <u>is approximately</u>

$N(\mu_E(t), \sigma_F^2(t))$ where

$\mu_E(t) = t(\tau + \mu_1)$ and   (=6.32t or 5 × 6.32t = 31.6t years)

$\sigma_E^2(t) = t\,\sigma_1^2$ (=2.898t)

<u>are respectively the mean and the variance of a normal</u>

<u>distribution</u>.

This theorem can **ex**plain all the stylized facts empirically observed (see Diagram 3). First of all, the theorem not only predicts that the relative frequencies of the generation birth schedules (b^{*t}) are bell shaped it also predicts that they are approximately normally distributed. Furthermore, they become more "regular" for later generations.

Next, the theorem states that the mean gap for two consecutive generations is $\mu_E(t+1) - \mu_E(t) = \tau + \mu_1$ or 31.6 years. Thus, the theorem not only predicts a constant mean gap but also tells us that the magnitude of this gap is the average fatherhood age. As we recall, the mean gap between the 11th, 12th and 13th generations of the Hsü clan are 31.6 and 31.4 years respectively. The mean gap between the 13th and the 14th generations is 29.5 years as computed from the observed frequency b^{*t}. Thus, the theoretical predictions are clearly borne out by the inductive evidences.

Finally the theorem predicts the diminishing concentration tendencies for E^i. Not only the birth range but also the variance increases as a function of generation. Thus the basic theorem gives plausible explanations to all the stylized facts for the Hsü clan -- as well as for many other family clans for which we are able to gather genealogical data.

Section 5: Multiple Ancestor Model

The single ancestor model depicts an idealized male regeneration process from a single ancestor. For the purpose of theoretical testing, the model is adequate only when the clan genealogy contains information that covers all generations. However, this is not typically the case. For the Hsü clan in our study, the single male ancestor was born in the fourteenth century. However, the genealogical information for this clan was very inadequate for all generations before the 11th generation (i.e., for the more than 300 years between 1300 and 1700).[11] Rather complete information is available only from the 11th generation on -- as we have presented in Table 1. When the data are incomplete for early generations -- which is the usual case -- the single ancestor model must be modified.

In a multiple ancestor model, a particular generation is first chosen as the ancestor generation while the birth schedule E^0 of this generation will be referred to as the ancestor birth schedule. E^0 is now a "schedule" containing all males of the same generation born in different years. All the males in F^0, who are in fact distant cousins, will be interpreted as "ancestors" and hence the name of a multiple ancestor model. When E^0 is given, a sequence of expected generation birth schedules E^1, E^2, E^3,... is generated when the same male fertility schedule is applied. The primary

empirical   data in this model is the pair (E^0, U^1), the ancestor birth

schedule and the male fertility schedule, while all other E^1, E^2, E^3,...

are deduced theoretically and recursively.

In applying this model for empirical research, the first generation

for which we have rather complete data on "male birth years" should be

treated as the ancestor generation. In the case of the Hsü clan, the 11th

generation is picked as the ancestor generation and hence $E^0 = b^0$ where b^0

is the empirically observed birth schedule of the <u>11th generation</u> (see Dia-

gram 3). Thus the primary data, in this case, is the pair (b^0, U^1), from

which a sequence of expected birth schedules E^1, E^2, E^3,... is to be

generated theoretically to be **compared** with (i.e., tested against) the empi-

rically observed schedules b^1, b^2, b^3,... corresponding to the 12th, 13th,

14th ... generations.

This theoretical procedure is illustrated in Diagram 3. For the 11th

generation (i.e., the ancestor generations) there is one curve $b^{*0} = E^{*0}$

which is the relative frequency of the ancestor birth schedule. For the

12th and 13th generations, there are <u>two</u> curves for each generation. In

addition to the observed sequence b^{*1}, b^{*2} (i.e., the solid curve), there

is an expected sequence E^{*1}, E^{*2} (i.e., the dotted curve) which is generated

theoretically (i.e., recursively).

Since the pair $(E^0 = b^0, U^1)$ provides all the demographic information

for the male regeneration process, in addition to the five indices defined

for U^1 in the last section, a number of indices will now be defined for

$E^0 = b^0$. First the sum of all entries in b^0 is $s_0^b = 284$ "persons", which is

the <u>total number of male ancestors</u>. The age difference between the oldest

and the youngest male in b^0 is $\gamma = 23$ (or $5 \times 23 = 115$ years) which is the

<u>ancestor birth range.</u> s_0^b has much to do with the future clan size while γ

affects the birth range for all future generations.

In the multiple ancestor model, it is natural to take the year in that the first ancestor was born as the 0-th year of the ancestor calendar. For the Hsü clan, the year 1665 is such a year and is the origin of the time axis in Diagram 3. To facilitate our exposition we shall refer to $E^{*0} = b^{*0}$ as a probability distribution function for an <u>ancestor random variable</u> x. In the ancestor calendar the mean value $\mu_0 = 11.22$ (or $5 \times 11.22 = 56.1$ years) of x is the <u>mean birth year</u> of the ancestors and $\sigma_0^2 = 19.37$ is the variance of b^{*0}. Thus, altogether there is a total of nine indices s_1, τ, n, μ_1, σ_1^2, s_0^b, γ, μ_0, and σ_0^2 defined for the pair (b^0, U^1) in the multiple ancestor model.

Notice that if $E^0 = b^0$ contains precisely one ancestor, the multiple ancestor model reduces to the special case of the single ancestor model in the last section. A direct generalization of the basic theorem in the last section (proved as corollary 2 in Appendix I) is:

<u>Generalized Theorem</u>: For large t, the relative frequency E^{*t} is describable approximately by the probability distribution function of a random variable $z^t = x + y^t$ where

(i) x is the ancestor random variable with mean μ_0 and variance σ_0^2 ,

(ii) y^t has a normal distribution $N(\mu_y, \sigma_y^2)$ where

$$\mu_y = t(\tau + \mu_1)$$
$$\sigma_y^2 = t\sigma_1^2$$

and hence the mean and variance of z^t are given by

(iii) $\mu_z(t) = \mu_0 + t(\tau + \mu_1)$ (mean : $\mu_z(t) = 11.22 + 6.32t$)

$\sigma_z^2(t) = \sigma_0^2 + t\sigma_1^2$ (variance : $\sigma_z^2(t) = 19.37 + 2.898t$)

This theorem which provides us with a method to approximate the recursively generated E^{*t} by a well define function $f(z^t)$, is illustrated in Diagram 4. For the 12th and the 13th generations the b^{*t} and E^{*t} curves of Diagram 3 are reproduced. For each generation, a third curve, corresponding to $f(z^t)$ is now added. Notice that E^{*t} and $f(z^t)$ are, for all purposes, indistinguishable. This nearly perfect approximation becomes more amazing when we realize that the above theorem predicts a good approximation only for large t regardless of the shape of U^1 or b^0. However, the near perfect approximation is seen to be already valid even for small t (i.e., for t=1 and t=2) corresponding to the 12th and the 13th generations.

The "immediacy" of the near perfect approximation is accounted for primarily by two factors. First, the male fertility schedule in U^1 is, itself, bell shaped. A typical father reaches the maximum fertility rate $u_3^1 = .436$ by the time he reaches the age group "30-34" containing the mean birth age ($\tau + \mu_1$). Thereafter the fertility rate declines steadily. Secondly, the ancestor birth schedule (or $E^{*0} = b^{*0}$) is also bell shaped. For the basic theorem of the single ancestor model already predicts that b^{*0} should be bell shaped, in fact, normally distributed. The fact that b^0 is slightly "out of shape" is due primarily to the fact that the genealogical data is rather incomplete for the 11th generation. The immediacy of the near perfect approximation of E^{*t} curve is due to the fact that the random variable z^t is affected by two bell shaped distributions (b^{*0} and U^{*1}).

The immediacy of the near perfect approximation of E^{*t} by $f(z^t)$ greatly simplifies the tasks involved in the testing of our theory. For instead of having to compare the observed b^{*t} with E^{*t}, all we have to do

is to compare b^{*t} with $f(z^t)$. Can $f(z^t)$ predict all the stylized facts? The explanation given so far already implies that b^{*t} is predicted to be bell shaped. However, unlike the single ancestor model, b^{*t} is now predicted to be a slightly distorted normal distribution defined by $f(z^t)$. The property of constant mean gap is also predicted to be $(\tau + \mu_1)$ or 31.6 years, which is exactly the same conclusion as we have reached for the single ancestor model. Finally, the theory also predicts the diminishing concentration property as the variance gap is predicted to be a constant (i.e., $\sigma_1^2 = 2.898$) as is the case for the single ancestor model. Notice that the ancestor birth schedule b^0 has nothing to do with the mean gap or the variance gap whatsoever.

The genealogical data for earlier generations (e.g., before the 11th generation for the Hsü clan) is often sparse. Theorems can help us to overcome some problems with the data scarcity. For example, for the Hsü clan, the genealogy only vaguely recorded that the first ancestor "was born at the end of the Yüan dynasty (1260-1367) and that he "moved to Hsiao-shan in the beginning of the Ming dynasty to avoid the population registration" which was ordered by the founding emperor in 1370.[12] We can predict his birth year by converting the mean equation $\mu_t = 11.22 + t6.32$ into Christian calendar to become $Y_t = 1665 + 5\mu_t$. Using this equation, his birth year is estimated to be 1342. Since he migrated to Hsiao-shan shortly after 1370, he must have migrated when he was slightly older than 28. The genealogy further recorded that "shortly after his arrival in Hsiao-shan he became an uxorilocal husband in a Wen family."[13] The age of 28+ which we predicted seems to be not far off.

The theorem (i.e., the equation for $f(z^t)$) can also help us to recapture the birth schedules of early generations (i.e., before the 11th generation for the Hsu clan) for which data is sparce. For example, in the hierarchy matrix in 1790 of Table 2, the birth schedule of the 10th generation (i.e., the first row) is, in fact, not available. However, by substituting t=-1 in the mean and variance equations in the basic theorem, the mean and variance are estimated to 4.9 and 16.47 respectively. We, therefore, can know at least the relative frequency of the birth schedule for the 10th generation because the basic theorem for the single ancestor model implies that it is normally distributed N(4.9, 16,47). We then only need to know the total male births in the 10th generation to approximate the birth schedule. (See theorem 2 in the next section).

Section 6: The Size of the Family Clan

In addition to theorizing on the pattern of the birth schedule in section 3 to 5, we will now theorize on the size of the family clan for which certain inductive evidences (e.g., the total male births in Diagram 1 and total male population in Diagram 2) have been examined in sections 1 and 2. As far as the "total male births" is concerned, the basic theorem allows us to construct an expected sequence of generation birth schedules E^1, E^2, E^3,.... When they are tabulated as in Table 1, the sum of all rows, denoted by $E^0 (=b^0) + E^1 + E^2 +...$ is the expected total male births and is shown by the dotted curve in Diagram 1. Notice that this dotted curve begins from 1710 and does not cover the entire generative phase (1665-1745) because the birth schedules before the 11th generation are not available. We can test our theory for the period from the year 1710 on.

We have a clearer idea of the demarcation of the generative phase from the transition phase by comparing the "observed" with the "expected" total male birth curves in Diagram 1. For the generative phase (more precisely, for approximately half a century from 1700 to 1750 for which the expected curve is available), the two curves follow the same time trend. In contrast, in the transition phase, the two curves clearly diverge as the observed total births grow at a much slower rate than the expected total births.

The above constrast between the two phases can be verified in another way via the use of a modern demographic concept. When the male fertility schedule and mortality schedule are given, we can derive what is called the stable population growth rate (or the "intrinsic" growth rate) by solving a well known equation.[14] Using the data of the Hsü clan (see Appendix II), the intrinsic growth rate is $r=.02$ as shown by the horizontal line in Diagram 2. It is that growth rate of the total clan population that prevails in the long-run. In the generative phase, the actual growth rate fluctuates around the intrinsic growth rate. In constrast, during the transition phase, the dotted curve drifts downward and away from the intrinsic growth rate. This clearly shows that the clan is beginning to "lose population" during the transition phase.

That the clan begins to lose population in the transition phase can be verified indirectly in still another way. Let s_0^b, s_1^b, s_2^b,... be the total male births of each generation (i.e., s_t^b is the sum of all elements in E^t). Obviously, s_t^b depends upon $s_0^b=284$ (i.e., the total number of ancestors in b^0) and $s_1=2.02$ (the total fertility of a typical father). In the multiple ancestor model, it is not difficult to prove (Lemma 3 of Appendix I):

Theorem 2: <u>The total male births of the t-th generation is given by</u>

$$s_t^b = s_0^b \, s_1^{b \, t}.$$

According to this formula, the expected total births are $(s_1^b, \; s_2^b, \; s_3^b)$ = (547, 1159, 2341) for the 12th, 13th, and 14th generations. When these numbers are used to multiply (i.e., to "blow up") the relative frequencies $(s_1^b E^{*1}, \; s_2^b E^{*2}, \; s_3^b E^{*3})$, the expected total birth schedules are represented by the bar diagrams super-imposed on those for the observed total births $(b^1, \; b^2, \; b^3)$ in Diagram 3. The amount that the latter fall short of (exceed) the former is represented by the shaded ("crossed") area.

For the 12th generation, except for the major catastrophic years (1745-1765) when major population losts are observed, the shaded areas roughly "cancel out" the crossed areas. For these years (i.e., before 1745 and after 1765), the fluctuations of the total births appear to be a chance (or probability) event. In contrast, for the 13th (and all later) generations, the shaded areas overwhelmingly dominate. Thus, the population losses cannot be explained merely as a chance event and that whatever reasons that have caused "underregistration" (i.e., the male births not recorded in the genealogy) must be of a consistent and persistent type, and not merely a "random" phenomenon.

That such non-random forces prevail and bring about a gradually disintegration of the family clan in the transition phase can be seen by comparing the expected total births with the observed total births for each generation: 284 (284), 547 (479), 1159 (651), 2341 (724). Here the observed total births for the 11th through 14th generations, indicated in the parenthesis, account for smaller fractions of the expected values in later generations (e.g., only 31% by the 14th generation).

Section 7: The Birth Range and the Structure of the Hierarchy Matrix

The disintegration of the family clan in the transition phase can be seen from still another way via the birth range of the expected generation birth schedules. Let f_t (similarly g_t) be the year of the first (last) birth of the t-th generation, then $r_t = g_t - f_t$ is the birth range (i.e., the maximum age difference between any two males of the same generation). In the multiple ancestor model, the following theorem is obvious:

Theorem 3: In the multiple ancestor model, the year of the first (last) birth, f_t (g_t), for the t-th generation is given by

$$f_t = Y_0 + t\,\tau \qquad (Y_0=1665, \quad \tau=3 \text{ or } 15 \text{ years})$$

$$g_t = Y_0 + \gamma + t(\tau + n-1) \quad (\gamma =115, \quad \tau+n-1 = 59 \text{ years})$$

where Y_0 is the birth year of the first male ancestor in Christian calendar. Hence, the birth range of the t-th generation is

$$r_t = g_t - f_t = \gamma + t(n-1).$$

For the Hsü clan, n=9 (or 9 × 5 = 45 years) and γ =23 (or 5 × 23 = 115 years) and hence r_t = 115 +40t. Using this equation, the birth range for the 11th, 12th, 13th, and 14th generations are expected to be 115 (115), 155 (135), 195 (140), and 235 (155) where the observed birth ranges are indicated in the parentheses. Again, we see that the observed birth ranges become smaller fractions of the expected values for later generations.

The number of rows (e.g., 9) in a hierarchy matrix (e.g., Table 2) indicates the number of consecutive generations N_a that coexist in a particular year (e.g., a=1790) in Christian calendar. As the birth ranges shrink from their expected values so does N_a. If $T_L(a)$ (similarly $T_F(a)$)

is the last generation (similarly, the first generation) that affects the hierarchy matrix of the a-th year, then $N_a = T_L(a) - T_F(a) + 1$. Using the notation $[x]$ to stand for the largest integer less than x (e.g., $[5.8] = 5$, $[-2.3] = -3$). the equations of f_t and g_t in theorem 3 lead immediately to the following theorem:

Theorem 4:   In the multiple ancestor model, with the ancestor generation identified as the 0-th generation, the last (first) generation $T_L(a)$ ($T_F(a)$) that has any survivors in the a-th year in Christian calendar is given by

$$T_L(a) = [(a-Y_0)/\tau] \qquad (Y_0=1665)$$

$$T_F(a) = [(a-L-Y_0-\gamma)/(\tau+n-1)] +1 \qquad (L=80)$$

where L is the male longevity, and hence the maximum number of generations that can coexist in the hierarchy matrix of the a-th year is $N_a = T_L(a) - T_F(a) + 1$.

In our case, a=1790, L=80, γ=115, τ=15, and n=45, we have T_F=-1 (i.e., the 10th generation) and T_L=8 (i.e., the 19th generation) as the 11th generation is identified with the 0-th generation. However, from Table 2, we can no longer observe any survivors in the four generations from the 16th to the 19th which, in principle, could have affected the hierarchy matrix in 1790. This is due primarily to the fact that in the year 1790, the family clan was well into the transition phase so that fewer than N_a=8+1+1=10 generations can coexist in the hierarchy matrix.

Section 8 : The Disintegration of the Family Clan

All evidence examined indicate that in the transition phase the family clan begins to disintegrate leading to its dissolution in the declining phase. What directly "explains" this disintegration is that the use of a constant male fertility schedule (U^1) is no longer adequate. Other institutional forces (i.e., forces other than purely "demographic" as summarized in U^1) begin to creep in. In this concluding section, we can only speculate on the nature of these other forces.

Two explanations for the disintegration may be rejected right away. One might conceivably attribute the disintegration as a purely demographic phenomenon brought about by a decreasing birth rate (i.e., a lower value for U^1). There is, however, no evidence to justify such secular change (described as "demographic transition" by the modern demographers) in this period. Another conceivable explanation is that there are missing data. However, this explanation also cannot stand on its own merit, as it is difficult to imagine that the very reliable data for the generative phase (i.e., that which corroborates our theoretical expectation) are to be replaced by a rather incomplete set for the later transition phase due to the negligence of the record keeper of the clan for a long stretch of time. In fact, the Hsü clan members did make such an effort to "complete the record" even after 1845.

The family clan, we believe, is a space sensitive social group, the cohesiveness of which must be analyzed in terms of distance and area. At any moment of time, a precondition for the performance of whatever clan functions (educational, ceremonial, economic, social security, etc.) is

that , relative to the rather primitive means of communication and transportation in a premodern society, the members (in the hierarchy matrix) must not lose contact with each other as they will when they live in such a way as to be separated by long distance. However, in an agrarian society this precondition cannot be fulfilled for long under population pressure, because of the shortage of land area for agricultural production. The family clan necessarily disintegrates under population pressure because emigration brought about by land area shortage renders allienation by distance unavoidable.

The county of Hsiao-shan, where the Hsü clan resided, is situated on a densely populated fertile plain southeast of Hangchow. Our conjecture is that the emigration of population from this region during the last quarter of the eighteenth century was the major economic cause of the disintegration of the Hsü clan.[15] The advantages offered by affiliation to the family clan constitute a very weak deterring force when the individual families decide to migrate for economic reasons. Furthermore, when the population size of the clan is too large, the scale diseconomy offers practically no resistance to disaffiliation. In other words, the record keepers did not make a serious attempt to register those who had migrated and scattered in distant lands in their periodic revisions of the clan genealogy.

Appendix I

Single Ancestor Model

    In this single ancestor model the family clan is started by a single male ancestor (i.e., the zeroth generation), the birth year of which is treated as the 0-th year of the ancestor calendar. Let $O_m = (0,0,\ldots 0)$ be a row vector with m-zeros, the male fertility schedule can be written as

Ala) $F^1 = (0_\tau, U^1)$ where

  b) $U^1 = (u^1_0, u^1_1, u^1_2, \ldots u^1_{n-1})$ with

  c) $s_1 = u^1_0 + u^1_1 + \ldots + u^1_{n-1}$

where τ is the non-productive period, n the fertility range, and s_1 the total number of sons expected to be born to a typical father during his life time.

    Let E^t be the birth schedule of the t-th generation, we have

A2a) $E^t = (0_{t\tau}, U^t)$ $t=1,2,\ldots$ where

  b) $U^t = (u^t_0, u^t_1, u^t_2, \ldots u^t_{t(n-1)})$ with

  c) $s_t = u^t_0 + u^t_1 + \ldots + u^t_{t(n-1)}.$

Notice that E^1 is defined in Al. For $t \geq 1$, E^t implies that the oldest (youngest) male is not expected to be born before (after) the $t\tau$-th $(t(\tau+n-1)+1)$ year of the ancestor calendar. Thus the birth range of E^t is $t(n-1)+1$ years, while the total births of the t-th generation is s_t. When a row vector $X_n = (x_1, x_2, \ldots x_n)$ is given, we can define the following matrix:

$$
A3)\quad R(m, X_n) = \begin{pmatrix}
x_1 & x_2 & x_3 & \cdots\cdots\cdots\cdots & x_n & 0 & 0 & 0 & \cdots\cdots & 0 \\
0 & x_1 & x_2 & x_3 \cdots\cdots\cdots & & x_n & 0 & 0 & \cdots\cdots & 0 \\
0 & 0 & x_1 & x_2 & x_3 \cdots\cdots\cdots & & x_n & 0 & \cdots\cdots & 0 \\
\vdots & & \ddots & & & & & \ddots & & \vdots \\
\vdots & & & \ddots & & & & & \ddots & 0 \\
0 & \cdots\cdots & 0 & x_1 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots & & & & & x_n
\end{pmatrix}
$$

$$m \times (m+n-1)$$

with m rows and m+n-1 columns. In this notation, U^{t+1} is related to (i.e., generated by) U^t as follows:

A4a) $U^{t+1} = U^t R^{(t)}$ where

  b) $R^{(t)} = R(t(n-1)+1 , U^1)$.

Notice that U^t has the same number of columns as $R^{(t-1)}$ which is $((t-1)(n-1)+1) + n - 1 = t(n-1)+1$, i.e., the birth range of F^t defined in A2b). We have the following lemma:

Lemma one: The sum of all elements in U^t (or E^t) is $s_t = s_1{}^t$ and the relative frequencies of U^t and F^t are, repectively

A5a) $U^{*t} = (1/s_1{}^t) U^1 R^{(1)} R^{(2)} \ldots\ldots R^{(t-1)}$

  b) $E^{*t} = (0_{t_\tau}, U^{*t})$.

Proof: A5b) follows from A2). Applying A4) recursively, we have
$$U^t = U^1 R^{(1)} R^{(2)} \ldots\ldots R^{(t-1)}.$$

Post multiply both sides by the column vector $(1,1,\ldots\ldots 1)$ we have $s_t = s_1{}^t$ by A1c) which implies A5a. QED

A random variable x with a probability distribution function $f(x)$ will be called "zero-rooted with a range n" if $f(x)=0$ except for $x=0,1,2,\ldots,n-1$. For these n-distinct values of x, $f(x)$ can be written in a vector form:

A6a) $P_x = (p_0, p_1, \ldots, p_{n-1})$ (i.e., $f(x) = P_x$ for $x=0,1,\ldots,n-1$)

  b) $p_0 + p_1 + \ldots + p_{n-1} = 1$

The mean μ_x and variance σ_x^2 of x can be unambiguously defined.

We can interpret E^{*t} and U^{*t} in A5b) as the probability distribution function of the zero rooted random variable z^t and S_t respectively with mean and variance denoted by $(\mu_E(t), \sigma_E^2(t))$ for z^t and (μ_t, σ_t^2) for S_t.

A5b) implies $z^t = S_t + t\tau$ and hence

A7a) $\mu_E(t) = \mu_t + t\tau$.

   b) $\sigma_E^2(t) = \sigma_t^2$.

For $t=1$, $\mu_E(1) = \mu_1 + \tau$ and σ_1^2 are, respectively, the mean birth year and

the variance of the male fertility schedule E^1. For $t \geq 1$, $\mu_E(t)$ and $\sigma_E^2(t)$

are, respectively, the mean and variance of the birth schedule of the t-th

generation in the ancestor calendar. We will need the following lemma:

Lemma 2:   If the zero rooted random variable x (similarly y) with a

   probability distribution function $P_x = (x_0, x_1, \ldots, x_{n-1})$

   (similarly, $P_y = (y_0, y_1, \ldots, y_{m-1})$ with a mean μ_x (μ_y) and

   variance σ_x^2 (σ_y^2), then z=x+y is a zero-rooted random variable

   with a probability distribution function:

A8a) $P_z = (z_0, z_1, \ldots z_{n+m-2}) = P_x R(n, P_y)$ with

   b) $\mu_z = \mu_x + \mu_y$ (mean of z) and

   c) $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ (variance of z).

Proof: A8bc) are obvious. Notice that z has a range $0,1,2,\ldots,n+m-2$ and

   is thus a zero rooted random variable. For any integer i satifying

   $0 \leq i \leq n+m-2$, $z_i = \Sigma x_k y_j$ summation over $k+j=i$. In case $m \leq n$, we

   have

$$z_i = \begin{cases} x_0 y_i + x_1 y_{i-1} + x_2 y_{i-2} + \ldots + x_i y_0 & \text{for } i \leq m \leq n \\ x_{i-n} y_m + x_{i-n+1} y_{m-1} + \ldots + x_i y_0 & \text{for } m < i \leq n \\ x_{i-n} y_m + x_{i-n+1} y_{m-1} + \ldots + x_{n-1} y_{i-(m-1)} & \text{for } m \leq n < i \end{cases}$$

In all cases, z_i is the i-th element in the row vector $P_x R(n, P_y)$.

For $n < m$, the proof is similar. QED

With the aid of this lemma, we can prove:

Theorem 1: If x_1, x_2,...x_t are independent and zero-rooted random samples from a population U^{*1}, the sample sum $S_t = x_1 + x_2 + \ldots + x_t$ is a zero-rooted random variable with a probability distribution function U^{*t} such that:

A9a) $\mu_t = t\,\mu_1$

b) $\sigma_t^2 = t\,\sigma_1^2$

Proof: A9ab) are obvious. To prove the theorem inductively, for t=1, $S_t = x_1$ and the theorem is true by definition. Suppose the theorem is proved for t, then $S_{t+1} = S_t + x_{t+1}$. The inductive hypothesis and lemma 2 imply S_{t+1} has a probability distribution function

$$U^{*t}R(t(n-1)+1, U^{*1}) = (1/s_1^{t+1})U^t R^t \quad \text{by A4b) and Lemma one}$$

$$= (1/s_1^{t+1})\,U^{t+1} \quad \text{by A4a)}$$

$$= U^{*t+1} \quad\quad\quad \text{by Lemma one.}$$

QED.

Substituting A9ab) in A7ab), we have:

A10a) $\mu_E(t) = t(\tau + \mu_1)$

b) $\sigma_E^2(t) = t\,\sigma_1^2$

where $\tau + \mu_1$ is the mean birth age of the male fertility schedule U^1.

The central limit theorem applied to lemma 2 leads to the following corollary:

Corollary 1: For large t, E^{*t} is approximately normally distributed $N(\mu_E(t), \sigma_E^2(t))$ where the mean $\mu_E(t)$ and variance $\sigma_E^2(t)$ are defined in A10ab.

This is the basic theorem of the single ancestor model.

Multiple Ancestor Model

For the multiple ancestor model, in addition to the male fertility schedule U^1 defined in A1), we need the ancestor birth schedule:

A11a) $b^0 = (b^0_0, b^0_1, b^0_2, \ldots b^0_{\gamma-1})$ with

  b) $s_0 = b^0_0 + b^0_1 + b^0_2 + \ldots + b^0_{\gamma-1}$

where s_0 is the total number of ancestors and where γ is the ancestor birth range. Let E^t (t=1,2,3,...) denotes the birth schedule of the t-th generation then

A12a) $E^t = (0_{t\tau}, b^t)$ where

  b) $b^t = (b^t_0, b^t_1, b^t_2, \ldots b^t_{t(n-1)+\gamma-1})$ with

  c) $s^b_t = b^t_0 + b^t_1 + \ldots + b^t_{t(n-1)+\gamma-1}.$

This is due to the fact that for the t-th generation the first (last) male is not expected to be born before (after) the $t\tau$-th year ($t(n-1)+\gamma-1$ year) in the ancestor calendar (i.e., the 0-th year is now the birth year of the first male in b^0). The sum of all elements in E^t (or b^t) is s^b_t, the total births of the t-th generation.

The b^0_i ancestor in b^0 generate a birth schedule $b^0_i U^t$ for the t-th generation. Since $b^0_{i+1} U^t$ lags behind $b^0_i U^t$ by one year, we have

A13) $b^t = b^0 R(\gamma, U^t).$

This leads to the following lemma:

__Lemma 3:__ The total male births (i.e., the sum of all elements in E^t or b^t) is $\underline{\text{is}}\ s^b_t = s_0 s_1{}^t$ (for s_1 defined in A1c) and the relative frequencies of b^* and E^* are

A14a) $b^{*t} = (1/s^b_t) b^0 R(\gamma, U^t) = b^{*0} R(\gamma, U^{*t})$

  b) $E^{*t} = (0_{t\tau}, b^{*t}).$

Proof: A14b) follows from A12a). Post mulitply both sides of A14) by a

column $(1,1,...1)$ leads directly to $s_t^b = s_0 s_1^t$ by lemma one (i.e.,

$s_t = s_1^t$). QED

This lemma involves four relative frequencies E^{*t}, b^{*t}, b^{*0}, and U^{*t}

which may be interpreted as the probability distribution functions of four

zero-rooted random variables z^t, w^t, x, and S_t with means and variances

denoted by $(\mu_E(t),\ \sigma_E^2(t))$, $(\mu_W(t),\ \sigma_W^2(t))$, $(\mu_0,\ \sigma_0^2)$, and $(\mu_t,\ \sigma_t^2)$.

A14b) implies that $z^t = w^t + t\tau$ and hence

A15a) $\quad \mu_E(t) = \mu_W(t) + t\tau$

b) $\quad \sigma_E^2(t) = \sigma_W^2(t)$

The relation between w^t, x and S_t is covered by the following theorem:

Theorem 2: If w^t, x and S_t are the zero-rooted random variable with

probability distribution functions b^{*t}, b^{*0}, and U^{*t}, then

$w^t = x + S_t$ and hence

A16a) $\quad \mu_W(t) = \mu_0 + \mu_t$

b) $\quad \sigma_W^2(t) = \sigma_0^2 + \sigma_t^2$

Proof: Apply lemma 2 to the pair "x and S_t" then A14a) implies the theorem

directly. QED

Substituting A16ab) in A15ab) and making use of A9a), we have:

A17a) $\quad \mu_E(t) = \mu_0 + t(\tau + \mu_1)$

b) $\quad \sigma_E^2(t) = \sigma_0^2 + t\sigma_1^2$

which are the mean and variance of E^{*t} of the t-th generation. Since

A18a) $\quad z^t = w^t + t\tau = x + S_t + t\tau \quad$ or

b) $\quad z^t = x + y$ where $y = S_t + t\tau$

the central limit theorem implies that, for large t, y is approximately

normally distributed. Thus we have

Corollary 2: For large t, the relative frequency E^{*t} is distributed as the sum of two random variable $z^t = x + y$ where x has the probability distribution function b^{*0} (i.e., relative frequency of the ancestor birth schedule with a mean μ_0 and variance σ_0^2) and where y is approximately normally distributed with a mean $t(T+\mu_1)$ and variance $t\sigma_1^2$. (The mean and variance of Z^t are defined in A17ab.)

This is the basic theorem of the multiple ancestor model.

Appendix II

In this appendix, we shall discuss the nature of data recorded in the genealogy of Hsü clan in Hsiao-shan (Hsiao-shan T'ang-wan Ching-t'ing Hsü-shih tsung-p'u 蕭山塘灣井亭徐氏宗譜 , published 1911, revised 1923) and the method of computing the empirical values of the birth schedules and male fertility schedule.

For the first three generations, the Hsü clan genealogy recorded only three names without information on vital dates. From the fourth generation on, the record of vital dates can be found at first spottily and gradually more completely. Table Al shows the number of males recorded by generations. These numbers and dates are the basic materials for our empirical evidence.

The generation birth schedules as shown in Table 1 of the text are derived by identifying the males in each generation according to their birth years at five-year intervals starting from 1663 (i.e., 1665 is the mid-point of the interval 1663-1667, etc.), the year recorded for the first male birth in the 11th generation. The males whose birth years are not known are not taken into consideration in the construction of the birth schedule. As can be seen from Table Al, from the 11th generation onward, the numbers with birth year known constitute well over 70 percent of the total births, thus, the deleted unknown numbers may not cause great bias in our observation.

Since the male fertility schedule (U^1) consists of two elements -- the male age-specific fertility rates and mortality rates -- each of them will be discussed separately.

The male age-specific fertility rates are computed by applying the technique of "family reconstitution" to the genealogical data. On each family reconstitution sheet, the vital dates of parents and sons are recorded and the age differences between the father and sons are calculated. The fathers are groups into five-year birth cohorts according to their birth years. The number of sons born to the fathers of the same cohort are distributed according to the ages of fathers in their reproductive period.

In computing the age-specific fertility rates, the numerator is the aggregate number of sons born to the fathers at each age group and the denominator is the aggregate person-years spent by the fathers at each age group. Thus, the age-specific fertility rates of the cohorts of fathers can be obtained. Furthermore, by taking nine consecutive cohorts into consideration at a suitable time point, the age-specific fertility rates as observed at this time point can be derived. Table A2 shows the male age-specific fertility rates computed from the data provided by the Hsü clan genealogy. In the text, the average value of the eight observed periods is used.

For computing the age-specific mortality rates, the technique of constructing a life table is applied to the genealogical data. Just as in the computation of the age-specific fertility rates, the males are first grouped into cohorts and then according to their age at death, the age distribution of deaths for each cohort can be obtained. The number of deaths at each age group can be added up from the last age group to the first age group to obtain the number of survivors at each age group. The ratios derived from dividing the number of deaths by the number of survivors at each age group are the age-specific death rates. The observed death rates

Table A1 The Hsü Clan Males Recorded in the Genealogy

Generations	Birth Year		Total	Ratio Known/Total	Year Known
	Known	Unknown			
4	1	3	4	.25	1458
5	1	7	8	.13	1505
6	4	27	31	.13	1528,1545,1540,1554.
7	7	55	62	.11	1552,1561,1572,1576,
					1582,1583,1593.
8	17	87	104	.16	1579-1639
9	42	119	161	.26	1605-1682
10	98	160	258	.38	1615-1735
11	284	122	406	.70	1663-1774
12	479	82	561	.85	1685-1816
13	651	54	705	.92	1721-1856
14	724	33	757	.96	1740-1890
15	704	43	747	.94	1773-1913
16	577	10	587	.98	1800-1913
17	303	3	306	.99	1835-1913
18	88	2	90	.93	1870-1913
19	5	0	5	1.00	1905-1913

Table A2 The Age-specific Fertility Rates of the Hsü Clan Male

Periods	Age Groups								
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
1725-1739	.0177	.0614	.0773	.0902	.1152	.0298	.0526	.0091	–
1740-1754	.0044	.0413	.0670	.0686	.1091	.0910	.0730	.0412	.0470
1755-1769	.0050	.0256	.0519	.0736	.0990	.0694	.0372	.0128	.0224
1770-1784	–	.0169	.0480	.0926	.0770	.0763	.0413	.0228	.0040
1785-1799	.0040	.0396	.0701	.0924	.0883	.0748	.0415	.0248	.0139
1800-1814	.0141	.0495	.0786	.0883	.0846	.0588	.0512	.0481	.0057
1815-1829	.0120	.0283	.0544	.0945	.0827	.0454	.0401	.0210	.0074
1830-1844	.0122	.0554	.0739	.1027	.0927	.0686	.0347	.0296	.0103
Average	.009	.074	.102	.122	.116	.079	.049	.028	.012

are graduated for the purpose of constructing a life table.

In addition to investigating the deaths by cohorts, the mortality can also be observed at a certain period by taking consecutive cohorts into consideration. Table A3 shows, for example, a life table constructed for the Hsü clan male in the period 1770-1784.

It should be noted that due to the nature of genealogical recording, the age-specific death rates of childern under age 15 cannot be derived directly from the data as the information of child deaths is usually very incomplete. Thus, for constructing a life table for all age groups, we have first constructed a life table of adults and then chosen two Princeton model life tables as references and extrapolated the observed values of death rates from the age of 15 up to age zero. This is perhaps not the best solution for the estimates of the child mortality. However, a more desirable method for estimating the child mortality from the incomplete data of Chinese genealogies has yet to be devised.

In the text, the survival rates (i.e., the l_x values in the life table) are taken from Table A3 to compute the male population for each year. Although in a preliminary study on the mortality based on the same genealogy, it has been found that the mortality seemed to increase during the nineteenth century, this set of survival rates is used throughout the period concerned in the text.

For obtaining the male fertility schedule, the average age-specific fertility rates as shown in Table A2 are multiplied by the survival rates at each age group and then times 5. (This is the same as to multiply the age-specific fertility rates by the life table values of L_x / l_x). In the text, we have used the values of L_x / l_x calculated for the Hsü clan male

cohort of 1695-1709. The values for the nine age groups of reproductive period are: 3.987, 3.805, 3.694, 3.573, 3.205, 2.921, 2.560, 2.128.

Table A3 Life Table of the Hsü Clan Male in 1770-1784

Age	q_x	l_x	d_x	L_x	T_x	e_x
0-1	.1337	10000	1337	9104	416321	41.63
1-4	.0788	8663	683	32843	407217	47.01
5-9	.0211	7980	168	39480	374374	46.91
10-14	.0152	7812	119	38763	334894	42.87
15-19	.0208	7693	160	38065	296131	38.49
20-24	.0295	7533	222	37110	248066	34.26
25-29	.0412	7311	302	35800	220956	30.22
30-34	.0570	7009	399	34048	185156	26.42
35-39	.0780	6610	516	31760	151108	22.86
40-44	.1055	6094	643	28863	119348	19.58
45-49	.1400	5451	768	25335	90485	16.60
50-54	.1863	4683	872	21235	65150	13.91
55-59	.2436	3311	929	16733	43915	11.52
60-64	.3150	2882	908	12140	27182	9.43
65-69	.4025	1974	795	7883	15042	7.62
70-74	.5086	1179	600	4395	7159	6.07
75-79	.6357	579	368	1975	2764	4.77
80+	1.0000	211	211	789	789	3.74

Notations:

q_x: Probability at age x of dying before reaching x+n.

l_x: Number of survivors at age x out of an original cohort of 10000.

d_x: Number of deaths between age x and x+n out of an original cohort of 10000.

L_x: Number of person-years lived between age x and x+n by an original cohort of 10000.

T_x: Number of person-years lived at age x and over by an original cohort of 10000.

e_x: Average number of years remaining to be lived (expectation of life) at age x.

Footnotes

1. See Maurice Freedman ed., <u>Family and Kinship in Chinese Society</u> (Stanford, 1970), for a comprehensive treatment on the subject of Chinese family.

2. See, for example, Hui-chen Wang Liu, <u>The Traditional Chinese Clan Rules</u> (Locust Valley, N. Y., 1959); Denis C. Twichette, "The Fan Clan's Charitable Estate, 1050-1760," in D. S. Nivison and A. F. Wright eds., <u>Confucianism in Action</u> (Stanford, 1959), pp. 97-133; Wolfram Eberhard, <u>Social Mobility in Traditional China</u> (Leiden, 1962); Ping-ti Ho, <u>The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911</u> (New York, 1962); and Evelyn S. Rawski, <u>Education and Population Literacy in Ch'ing China</u> (Ann Arbor, 1979).

3. The growth of the clan population in the declining phase cannot approach the theoretically deduced "stable population growth rate", a key demographic concept, even in the long-run. Thus the genealogical data tend to be "unreliable" if viewed from a purely demographic standpoint. A positive view toward the reliablity of Chinese genealogies, see T. H. Hollingsworth, <u>Historical Demography</u> (Ithaca, 1969), pp. 200-201.

4. Typically, the basic information provided by Chinese genealogies is the birth and death dates of males and their spouses as well as how many sons they had. The procedure for the construction of the birth schedules will be briefly described in Appendix II.

5. It will be shown that in fact in the year 1790 precisely ten generations can coexist although the survivors are not observed in the 16th through 19th generations for a reason that will be explained later.

6. The computation of age-specific survival rates is discussed in Appendix II. In the last row of Table 2, we indicate the "directly observed" survivors in 1790 based on the information of death dates of individuals involved. Notice that the total number of survivors indicated in this row is 844 which is larger than the number 702 derived above. The difference between the two is due primarily to the fact that while birth dates are available in the genealogy for all the 1232 males (excluding those in the 10th generation), the death dates of only 844 of them are given. It is for this reason of incomplete information on death dates that we have employed the survival rates in Table 2.

7. Notice that this curve begins from the year 1700 because we have not made use of the generation birth schedules before the eighth generation. Even for the eighth, ninth, and tenth generations, we have derived the birth schedules by an approximate estimation of extrapolation.

8. The ratio of the annual births to the population is the crude birth rate. In our case, the crude birth rate in terms of male only varies between 14 and 44 per 1,000 (see row 12 in Table 1) testifying to the reasonableness of the data.

9. See, Hsiao-shan hsien-chih kao (The draft gazetteer of Hsiao-shan, 1935 ed.), 5:26-31, for the chronology of natural calamities occurred during the Ch'ing period. The two quotations appear in 5:27 and 5:29 respectively.

10. For the concept of male fertility schedule (i.e., the net reproductive rate), see Henry S. Shryock et al, Methods and Materials of Demography (Washington D. C., 1971), p. 541. What lies behind U^1 are the age-specific fertility rates and survival rates, see Appendix II for the calculation.

11. The reason for beginning with the 11th generation is due primarily to the fact that the birth dates are incomplete for the earlier generations, see Appendix II.

12. See Ping-ti Ho, <u>Studies on the Population of China, 1368-1953</u> (Cambridge, Mass., 1959), pp. 4-7 for a discussion on the population registration in the early Ming period.

13. <u>Hsiao-shan T'ang-wan Ching-t'ing Hsü-shih tsung-p'u</u> (The genealogy of Hsü clan), <u>chüan</u> 3.

14. The equation is $\int_{o}^{\omega} e^{-ra} p(a)m(a)da = 1$. See Ansely J. Coale and Paul Demeny, <u>Regional Model Life Tables and Stable Populations</u> (Princeton, 1960), pp. 9-10 for a brief discussion on Lotka's original idea of stable population; also see Henry S. Shryock, <u>Methods and Materials of Demography</u>, p. 528 for the method of calculation.

15. While a detailed study on the problem of migration has not yet available, for a brief discussion on the phenomenon of interregional migrations during the Ming and Ch'ing periods, see Ping-ti Ho, <u>Studies on the Population of China</u>, ch. VII.

Table 1 The Generation Birth Schedules, Estimated Population
and Births, Male, Hstd Clan in Hsiao-shan

Row	Gen.	1665	1670	1675	1680	1685	1690	1695	1700	1705	1710	1715	1720	1725	1730	1735	1740	1745	1750	1755	1760
(1)	11	2	3	5	5	10	6	11	17	17	16	25	31	18	24	17	25	26	8	10	3
(2)	12				2		4	2	4	5	9	19	20	17	23	39	35	42	29	35	22
(3)	13												3	3	6	13	18	25	11	31	35
(4)	14															1	1	1	0	8	9
(5)	15																				
(6)	16																				
(7)	17																				
(8)	18																				
(9)	19																				
(10)	Total Births	16	17	21	22	28	17	27	33	33	33	50	59	41	55	70	79	94	48	84	69
(11)	*Estimated Population								175	192	206	236	270	287	314	353	397	448	459	500	525
(12)	Birth rate per 1000									37	34	32	42	44	35	40	40	42	21	34	26

* In the period 1665-1730, the number includes the births estimated for the ninth and tenth generations.

Table 1 (continue)

1765	1770	1775	1780	1785	1790	1795	1800	1805	1810	1815	1820	1825	1830	1835	1840	1845	1850	1855	1860	1865	1870
3	1	1																			
23	31	30	20	20	19	13	9	4	1	2											
28	45	33	50	47	50	38	43	35	25	29	26	16	13	8	10	6	2	2			
8	5	20	26	27	27	52	45	60	61	68	57	32	41	39	31	34	13	20	9	4	9
		2	3	4	9	11	13	25	24	22	35	34	55	54	74	73	44	51	20	18	21
						1	1	3	4	6	10	8	11	22	27	31	35	32	36	45	27
														1	4	9	3	8	6	5	15
																			2		
62	82	86	99	98	105	114	111	127	115	127	128	90	120	124	146	153	97	113	71	72	74
542	572	599	638	669	702	741	773	817	847	884	925	925	953	977	1019	1061	1047	1051	1014	978	942
23	29	29	31	29	30	31	29	31	27	29	28	19	25	25	19	19	19	22	14	15	16

Table 1 (continue)

1875	1880	1885	1890	1895	1900	1905	1910	Total
								284
								479
								651
8	2	5	2					724
19	21	15	15	16	10	7	8	703
42	53	46	42	29	21	29	17	577
16	22	26	35	33	36	54	30	303
2	2	6	11	11	11	27	16	88
						4	1	5
87	100	98	105	89	78	121	72	
920	906	892						
19	22	22						

Diagram 1

Number of Births

Generative Phase ——— Transition Phase ——— Declining Phase

Exogenous Disturbance

Exogenous Disturbance

Expected births computed by generation ($b^0 + F^1 + F^2 + \dots$)

Observed total births ($b^0 + b^1 + b^2 + \dots$)

flood & epidemic

bad harvest

flood

Major famine

bad harvest

flood

typhoon

bad harvest

flood and famine

Major drought & flood

famine

flood

Taiping Rebellion

300 250 200 150 100 50 0

1665 1675 1685 1695 1705 1715 1725 1735 1745 1755 1765 1775 1785 1795 1805 1815 1825 1835 1845 1855 1865 1875 1885 1895 1905

Time

Table 2   The Hst'd Clan Hierarchy Matrix in 1790

Age	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	5	1	Total
Gen. \ Time	1710	1715	1720	1725	1730	1735	1740	1745	1750	1755	1760	1765	1770	1775	1780	1785	1790	
10	8 (0) G	6 (0)	5 (1)	3 (1)	2 (1)													24 (3)
11	16 (1)	25 (2)	31 (4)	18 (4)	24 (7)	17 (6)	25 (12)	26 (14)	8 (5)	10 (7)	3 (2)	3 (2)	1 (1)	1 (1)				242 (67)
12	9 (0)	19 (1)	20 (2)	17 (3)	23 (7)	39 (15)	35 (17)	42 (23)	29 (18)	35 (23)	22 (15)	23 (17)	31 (23)	30 (23)	20 (16)	20 (16)	19 (16)	442 (235)
13	H		3 (0)	3 (1)	6 (1)	13 (5)	18 (8)	25 (14)	11 A (7)	31 (21)	35 (25)	28 (20)	45 (34)	33 (25)	50 (39)	47 (37)	50 B (43)	398 (280)
14							1 (0)	1 (0)	0 (0)	E 8 (5)	9 (6)	8 (6)	5 (4)	20 (15)	26 (20)	27 (22)	27 (24)	132 (102)
15														2 (2)	3 (3)	4 (3)	9 (8)	18 (15)
16									D								0 C	0
Total	33 (1)	50 (3)	59 (7)	41 (9)	55 (16)	69 (26)	79 (37)	94 (51)	48 (29)	84 (56)	69 (48)	62 (45)	82 (62)	86 (66)	99 (77)	98 (78)	105 (91)	1256 (702)
Survival Rates	.0211	.0579	.1179	.1974	.2881	.3811	.4683	.5451	.6094	.6610	.7009	.7311	.7533	.7693	.7812	.7980	.8663	
Observed Survivors	0	3	6	5	16	27	48	62	30	70	51	58	80	86	99	98	105	844
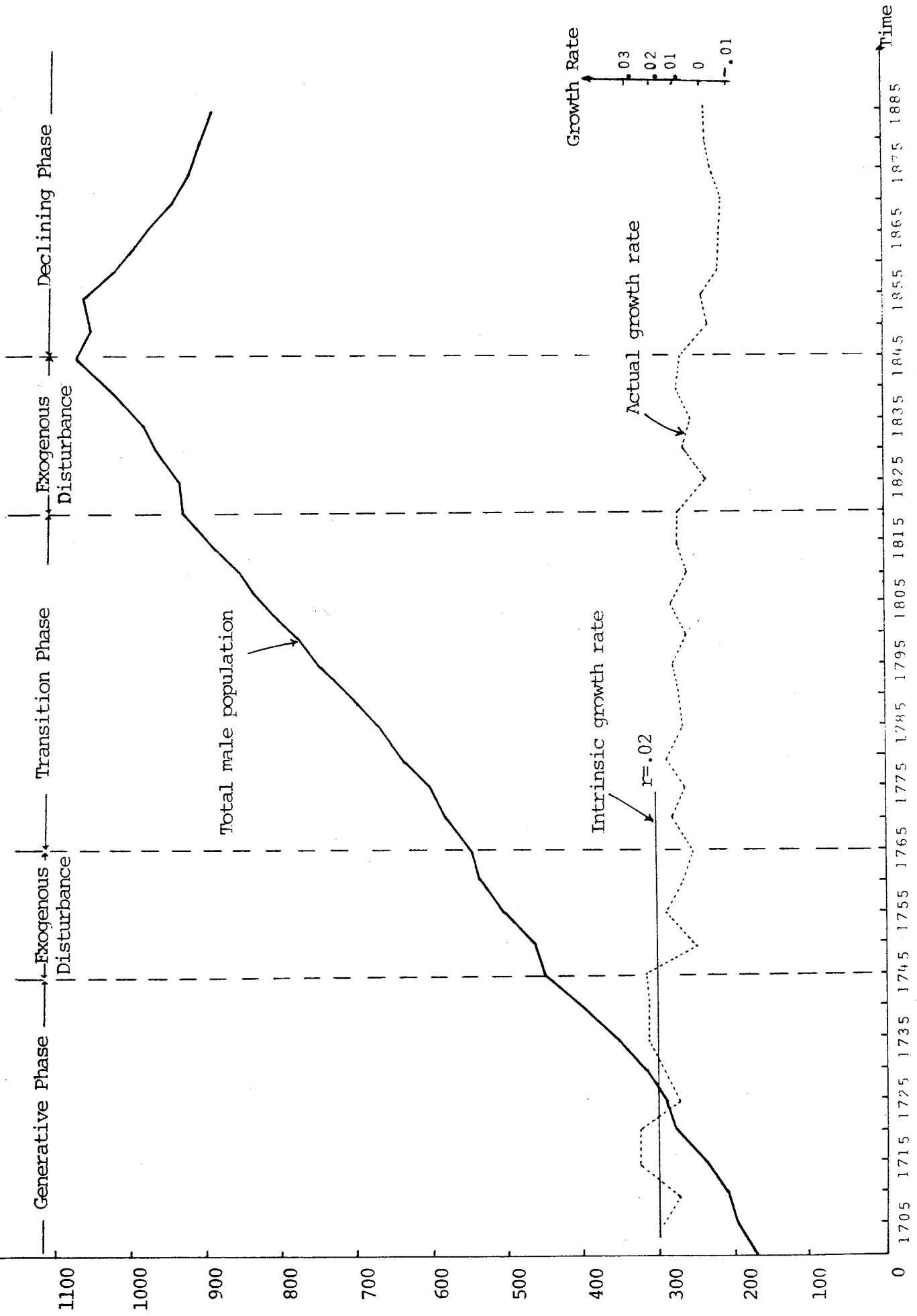
F

Diagram 2

Diagram 3

Diagram 4

Relative Distribution

Thirteenth Generation

$z^{(2)} = x + y^{(2)}$

F_i^{*2}: Expected relative frequency

b^{*2}: Observed relative frequency

$\tau + \mu_1^2 = 6.32$

$\mu^2 = 23.86$

Twelfth Generation

$z^{(1)} = x + y^{(1)}$

E^{*1}: Expected relative frequency

b^{*1}: Observed relative frequency

$\mu^1 = 17.54$

Time

Time